



COORDINATED PRODUCTION
FOR BETTER RESOURCE EFFICIENCY

D2.6 Report on automatic rule extraction methods and results

Lead Author: INEOS

Involved Partners: INEOS, Divis

May 2019

www.spire2030.eu/copro



Deliverable 2.6

Report on automatic rule extraction methods and results

Project Details

PROJECT TITLE	Improved energy and resource efficiency by better coordination of production in the process industries
PROJECT ACRONYM	CoPro
GRANT AGREEMENT NO	723575
INSTRUMENT	RESEARCH AND INNOVATION ACTION
CALL	H2020-SPIRE-02-2016
STARTING DATE OF PROJECT	NOVEMBER, 1ST 2016
PROJECT DURATION	42 MONTHS
PROJECT COORDINATOR (ORGANIZATION)	PROF. SEBASTIAN ENGELL (TUDO)

THE COPRO PROJECT

The goal of CoPro is to develop and to demonstrate methods and tools for process monitoring and optimal dynamic planning, scheduling and control of plants, industrial sites and clusters under dynamic market conditions. CoPro pays special attention to the role of operators and managers in plant-wide control solutions and to the deployment of advanced solutions in industrial sites with a heterogeneous IT environment. As the effort required for the development and maintenance of accurate plant models is the bottleneck for the development and long-term operation of advanced control and scheduling solutions, CoPro will develop methods for efficient modelling and for model quality monitoring and model adaption.

The CoPro Consortium

Participant No	Participant organisation name	Country	Organisation
1 (Coordinator)	Technische Universität Dortmund (TUDO)	DE	HES
2	INEOS Manufacturing Deutschland GmbH (INEOS)	DE	IND
3	Covestro Deutschland AG (COV)	DE	IND
4	Procter & Gamble Services Company NV (P&G)	BE	IND
5	Lenzing Aktiengesellschaft (LENZING)	AU	IND
6	Frinsa del Noroeste S.A. (Frinsa)	ES	IND
7	Universidad de Valladolid (UVA)	ES	HES
8	École Polytechnique Fédérale de Lausanne (EPFL)	CH	HES
9	Ethniko Kentro Erevnas Kai Technologikis Anaptyxis (CERTH)	GR	RES
10	IIM-CSIC (CSIC)	ES	RES
11	LeiKon GmbH (LEIKON)	DE	SME
12	Process Systems Enterprise LTD (PSE)	UK	SME
13	Divis Intelligent Solutions GmbH (divis)	DE	SME
14	Argent & Waugh Ltd. (Sabisu)	UK	SME
15	ASM Soft S.L (ASM)	ES	SME
16	ORSOFT GmbH (ORS)	DE	SME
17	Inno TSD (inno)	FR	SME

Deliverable 2.6

Report on automatic rule extraction methods and results

Document details

DELIVERABLE TYPE	REPORT	
DELIVERABLE NO	2.6	
DELIVERABLE TITLE	REPORT ON AUTOMATIC RULE EXTRACTION METHODS AND RESULTS	
NAME OF LEAD PARTNER FOR THIS DELIVERABLE	INEOS MANUFACTURING DEUTSCHLAND GMBH	
LIST OF AUTHORS - NAME(S) AND ORGANISATION(S)	PATRICK DENNIS SCHIERMOCH (INEOS), SINA DÄUBENER (DIVIS)	
VERSION	0.5	
CONTRACTUAL DELIVERY DATE	30 APRIL 2019	
ACTUAL DELIVERY DATE	09 MAY 2019	
Dissemination level		
PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission)	

Abstract

In large parts of the process industry process data is readily available and an increase of computational power in the recent years has raised attention for the use of this data for the purpose of modelling and control. A widely used concept is the black box modelling that uses process data to describe processes. It is especially advantageous in terms of the development time when compared to first principle models. However, a major drawback of black-box models is that the causality is lost due to the fact that multiple effects are lumped into each other to describe a certain aspect of a process.

In this report methods to connect causality to these black-box models in a target oriented way are described. One methodology deals with the deviations of current operation points to predicted values. Historical data of resource efficient operation points similar to the current operation is used to give plant operators hints how to improve. The second methodology that is discussed compares resource efficient and inefficient historical operation points to find rules that caused inefficient operation in the past. These will then be used to prevent inefficient production caused by the same reasons. Both of these concepts were implemented in real process plants in the chemical industry and examples are shown where reductions of resource consumptions by 40% and 10% are possible. Finally the results are discussed and further improvements are suggested to increase the degree of automation of the methodologies.

REVISION HISTORY

The following table describes the main changes done in the document since it was created.

Revision	Date	Description	Author (Organisation)
V0.1	27 February 2019	Creation	Patrick Dennis Schiermoch (INEOS)
V0.2	18 March 2019	Addition of the Section for CART-analysis and its implementation for INEOS in Cologne	Sina Däubener (Divis)
V0.3	22 March 2019	Implementation of first Revisions	Patrick Dennis Schiermoch

Deliverable 2.6

Report on automatic rule extraction methods and results

			(INEOS)
V0.4	30 March 2019	Review	Christian Jasch (Lenzing)
V0.5	03.04.2019	Implementation of the corrections from the reviewer	Patrick Dennis Schiermoch (INEOS)
V0.5	09 May 2019	Final approval	Sebastian Engell (TUDO)

Disclaimer

THIS DOCUMENT IS PROVIDED "AS IS" WITH NO WARRANTIES WHATSOEVER, INCLUDING ANY WARRANTY OF MERCHANTABILITY, NONINFRINGEMENT, FITNESS FOR ANY PARTICULAR PURPOSE, OR ANY WARRANTY OTHERWISE ARISING OUT OF ANY PROPOSAL, SPECIFICATION OR SAMPLE. Any liability, including liability for infringement of any proprietary rights, relating to use of information in this document is disclaimed. No license, express or implied, by estoppels or otherwise, to any intellectual property rights are granted herein. The members of the project CoPro do not accept any liability for actions or omissions of CoPro members or third parties and disclaims any obligation to enforce the use of this document. This document is subject to change without notice.

Table of contents

1	Executive summary	6
2	Introduction	7
3	Automatic rule extraction	8
3.1	CART- Analysis	9
4	Use Cases	10
4.1	INEOS in Cologne use cases	10
4.1.1	Root-cause analysis at INEOS in Cologne	11
4.1.2	Divis approach to automatic rule extraction at INEOS in Cologne.....	14
5	Results and discussion	15
5.1	Discussion of the Root-Cause analysis	15
5.2	Discussion of the CART-analysis	17
6	Summary and outlook	19
7	References	20

1 Executive summary

The CoPro project researches methods to increase the resource efficiency of the process industry by better coordination. One part of the project deals with the efficient development of models for process plants. A common practice is the development of black-box models that usually use historical data to describe a certain aspect of a process. The major advantage is that the effort is reduced when compared to first principle models. However, a major drawback is that the interpretability of changes in the results of the model is often heavily reduced.

In this report methods are described that use measured data from process units in order to connect causality with the deviations of real process data to the predicted values from models. The main goal is to guide plant operators to a more resource efficient production. Furthermore the acceptance of the previously mentioned black-box models is increased if they are not perceived as unrealistic.

The first concept described in this report involves the comparison of resource efficient production points close to the predicted values of the models with the current operation points. The goal of this method is to give plant operators useful hints for a more resource efficient operation.

The second concept compares historical data that is resource efficient with inefficient production points and returns general rules for the process plant that describe reasons for inefficient production in the past. These can then be used to avoid these inefficient operation points in the future.

It is shown that the implementation of both of these concepts in real process plants is able to increase the resource efficiency.

2 Introduction

The process industry in Europe is a highly competitive market that pressures its companies to innovative ideas if they want to compete in the future. One approach is to decrease the running costs by continuously increasing the resource efficiency of their processes. Furthermore, ambitious climatic goals were set and are enforced by political bodies that call for a more sustainable production throughout Europe.

To deal with the ambitious political requirements and with the rising significance of terms like industry 4.0 and digitalization a method to describe these processes in a computer understandable way becomes necessary. It is usually done by mathematical models that are varying in accuracy and detail. However, the continuous progress of the process industry and the coupling of sites due to economic reasons lead to higher degrees of complexity. Therefore, first principle models able to describe all of these effects to attain a certain degree of accuracy are needed. This is a high hurdle for the development and use of process modelling in the industry, which challenges researchers of many fields to find efficient approaches to process modelling. Furthermore, the development of models is also hindered due to the fact that some effects are not even fully understood yet. This circumstance is sparking the interest in alternative methods to describe processes and accumulate knowledge about them in novel and efficient ways.

The process industry is highly automated and collects massive amounts of process data. While this source of information has usually been neglected in the past the progress in computer technology enabled the use of this data to derive data-driven methods for process modelling, control and monitoring. It resulted in an increased interest for data based approaches in recent years. Data based model approaches can simplify the development of process models. However, they are often input-output models and relations between important factors are lumped together and cannot be differentiated anymore, leading to new challenges when it comes to the interpretability of the results. These models are referred to as black-box-models, because of the non-transparent nature of the changes to the results for different inputs to the model. To interpret the results and derive the necessary steps to steer the production into the desired direction extensive process knowledge and experience by operation personnel is necessary. Black-box models range from simple models that use low dimensional data inputs to describe the physical relations of a process to a wide variety of models that use high dimensional data to describe complex processes. Complicated relations can be described by these models and even with extensive process knowledge it is difficult to interpret the results. There is a need to come up with novel ways to connect causality with the benefits of black-box-modelling of complex processes to increase their transparency. This also results in a higher acceptance of the developed models because the user is supplied with a reason for the changes of the results.

The approaches presented in this report were developed during the framework of the CoPro project. In the first one rules for plant operators are automatically generated based on the current operation point. It is compared to historical data close to the values of a black-box baseline model that indicates the amount of resource consumption for resource efficient operation similar to the current operation. The goal is to facilitate the transition of the current operation to a more resource efficient operation point. Another use to automatic rule extraction is presented, where historic data is used to find heuristic guidelines for the plant operators to prevent the inefficient use of resources.

3 Automatic rule extraction

During the European project MORE multiple Key Performance Indicators (KPI) were defined and later used for the development of process models. In the CoPro project a tool was developed that expanded on this idea and uses surrogate models as well as process data to derive black-box models for baselines (B. Beisheim, 2018). These baselines give a reference point for the resource efficient resource consumption of the current operation in chemical plants based on historical plant data. The developed models are used to monitor the resource consumption of the process plants in real time. By giving the plant operators a reference point the processes are operated with higher resource efficiencies. However, while the indication of these resource consumptions is helpful, reaching these resource efficient operation points in the process plants

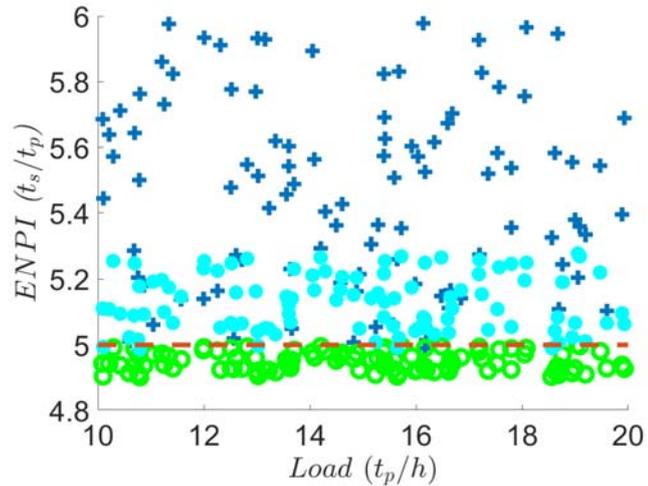


Figure 1: Schematic representation of the deviations of the resource consumption for different qualities of process control (green=APC, dark blue= operator, light blue=Automatic rule extraction)

relies heavily on the experience and knowledge of the plant operator because the reasons for the deviations in the current operation are not given by this black-box model. To further discuss this concept it is visualized in Figure 1. The baseline is indicated as a red stroked line. The EnPI on the y-axis is the specific consumption of a resource (e.g. steam) per ton of product leaving the plant, which is one of the previously mentioned KPIs. The load of the plant is given on the x-axis, which is used as input for the black-box model. The model inputs are usually influencing factors that are not controllable by the plant operator and affect the plants resource consumption. Figure 1 can be used to demonstrate the dependency between the EnPI and the influencing factor, however, in this case the EnPI is chosen to be constant to explain the concept of the baselines on a simplified example.

To reach the baseline different approaches can be pursued. The dark blue crosses indicate one end of the spectrum. Average plant operators operate the plant in different areas of resource efficiency ranging from high resource consumptions (120% of the baseline value) to very efficient values. Deviations like this are caused by different effects. Well trained operators are more likely to achieve performances closer to the baseline, while less experienced operators are prone to operate the plant inefficiently resulting in higher resource consumptions. Furthermore, independently from the experience of the operators it is not possible for a human to react to changes in the plant instantaneously. This delay results in inefficient plant operation and therefore deviations from the baseline.

On the other end of this spectrum there are highly automated plants that can use models to operate in resource efficient regions at all times. This is indicated by the green circles in Figure 1. The implementation of full automatic control in process plants is quite costly and requires complex control structures. However, it is very effective in maintaining an efficient operation.

An intermediate step to increase the use of the operational improvement potential of baseline models is the use of automatic rule extraction in the process plants in combination with the baselines for the resource efficient production. The general idea is to use historical plant data to derive reasons for the deviations from the efficient operation points. The implementation of the automatic rule extraction would ideally result in a closer distribution of the resource consumptions, which is indicated by the light blue dots in Figure 1. In the CoPro project two different approaches for automatic rule extraction have been tested, where one compares the current operation with operation points in the past under similar plant conditions and better resource efficiency. The reasons for deviations are then shown to the plant operator to support them with the identification of actions in order to increase the resource efficiency. These actions will ultimately result in lesser deviations from the baseline when compared to unassisted operation. The second approach uses a classification of the historical data to develop heuristic rules for the operation of the plants. These can then be used for process monitoring to avoid the inefficient operation of process plants. The theoretical background for both of these approaches is elaborated in the upcoming sections.

3.1 CART- Analysis

Classification and regression trees (CART) (L. Breiman, 1984) are a simple yet powerful tool to perform nonlinear regression and classification without using a black box model. The interpretability of this approach makes it especially powerful when it comes to understanding the structure and dependencies of a data set. Hence, this approach is used to identify heuristic rules for the root cause analysis in the process plants at INEOS in Cologne. In this section the theoretical background of the CART-Algorithm for classification is described. The general idea, visualized in Figure 2, is to split a

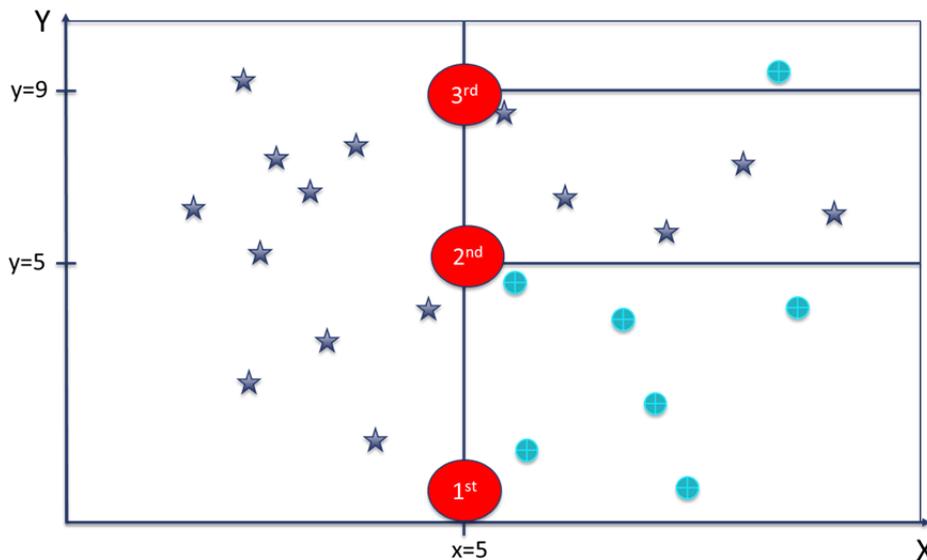


Figure 2: CART procedure on a 2-dimensional 2-class classification problem. The first data separation takes place at $x=5$. Points which have a value lower than 5 are classified as stars. The dataset with $x>5$ is again divided at $y=5$. The dataset for $x>5$ and $y<5$ is

data set based on a variable combined with an optimal threshold, so that the data in the resulting two split sets are more homogeneous than the combined overall dataset.

This idea can be mathematically expressed by the following: Let $x_i = (x_{i1}, \dots, x_{ip})$ with $i \in \{1, \dots, n\}$ be one observation vector containing information about p variables. y_i with $i \in \{1, \dots, n\}$ takes a value in $\{1, \dots, C\}$ and is the corresponding class to the input vector x_i . If the goal is to split the data

in a way that classes are separated optimally, a criterion for the impurity of the class is needed. The Gini impurity is used for this task in CART which is given by $GI(G) = 1 - \sum_{c=1}^C p(c)^2$ with $p(c)$ being the probability of classifying a random point in a subset G as class c . This can be empirically calculated through $p(c) = \frac{n_c}{m}$ with m being the cardinality of subset G and n_c being the number of observations in G with class label c . Thus, the classification problem can be expressed as an optimization problem:

$$\text{information gain: } \max_{j,t} (GI(G) - GI(G_1) - GI(G_2))$$

with $G_1 = \{(x_i, y_i) \in G \mid x_{ij} \leq t\}$ and $G_2 = \{(x_i, y_i) \in G \mid x_{ij} > t\}$.

This optimization can be continued iteratively using G_1 and G_2 as the new data set that should be split until a termination criterion is fulfilled or each class is successfully separated. Termination criterions can for example be defined as the information gain undercutting a certain threshold value or by setting a minimal number of observations that are required in a subset, also referred to as node. The resulting set of splitting rules can be presented in a tree structure as shown in Figure 3.

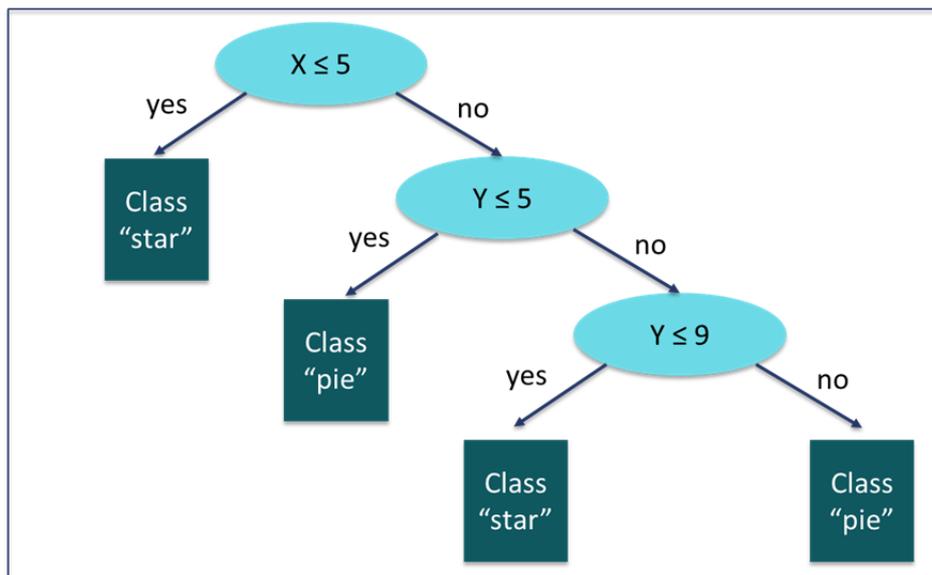


Figure 3: Classification tree derived from Figure 2

4 Use Cases

In the following section the use cases are discussed that were implemented at INEOS in Cologne for the automatic rule extraction. The previously described methods were used on multiple plants in order to increase their resource efficiency.

4.1 INEOS in Cologne use cases

INEOS is a company in the chemical sector that consists of a multitude of groups. The INEOS site in Cologne is a coupled industrial production site producing various chemical products in a wide range of different process plants. To circumvent the high effort that is bound to the development of first principle models for the numerous plants at this site a black-box baseline model is used to identify

resource efficient plant operation under the consideration of uninfluenceable process conditions. The implementation of this baseline concept for process plants at INEOS in Cologne is able to reduce the resource consumption in the production plants significantly.

However, the maximal improvement potential indicated by the baseline can rarely be utilized in everyday operation because deviations of measurements from the optimal conditions can be subtle and difficult to identify. Another aspect that hinders the achieving of a more resource efficient production is a conservative approach to changes in the plant conditions to maintain the product quality. Complex interdependencies result in a limited understanding of the effects that are invoked by alternating process parameters. This aversion to interventions in the process flow is lower with experienced plant operators but ultimately results in suboptimal process performance. Some plant operators are prone to the use of comfort values that decrease the plants efficiency in exchange of a lower effort to maintain the product quality. This behavior is not exclusively caused by the negligence of the plant operator but also by the fact that multiple process units have to be monitored and operated at once by a single operator, which can be overwhelming. Automatic rule extraction can help the plant operators to find appropriate values or prevent inefficient ones in order to streamline the plant operation resulting in a more resource efficient plant operation and a higher degree of comfort while operating the plant. Furthermore, the acceptance of black-box models can be increased if the results are not perceived as unreachable.

4.1.1 Root-cause analysis at INEOS in Cologne

At INEOS in Cologne a method has been developed to analyze data of the current operation point of a plant and compare it to historical data for similar process conditions in order to identify reasons for deviations from a resource efficient baseline. The procedure can be divided into the following nine steps:

1. Definition of a reference value with desirable process conditions
2. Determination of allowed deviations from these desirable process conditions
3. Acquisition of historical plant data for the process unit
4. Definition of a neighborhood of historical data that is similar to the current operation point
5. Definition of a “golden batch” of data that represents desirable process conditions similar to current process conditions
6. Standardization and normalization of the “golden batch” measurement data
7. Calculation of the standardized deviation of measurements from the current operation to the standardized “golden batch” measurements
8. Definition of trigger thresholds
9. Analysis of deviations beyond the trigger threshold

In the following part of the report these nine steps and their implementation at INEOS in Cologne are further elaborated.

In the first step a method to identify efficient process conditions is mandatory. At INEOS in Cologne black-box models are implemented for multiple process plants to estimate the Best Demonstrated Practice (BDP) of resource consumption using historical data. It is used to provide plant operators with a reference value to the amount of resource consumption of the plant considering the current process conditions. These models are designed in a way to return the specific amount of a resource

Deliverable 2.6

Report on automatic rule extraction methods and results

that should be consumed in consideration of conditions in the process plant that are uninfluenceable by the plant operator.

However, it is highly unlikely that these returned reference values and the actual resource consumptions represented in the historical data are going to be exactly the same and efficient operation points can be evaluated this way. To circumvent this limitation a certain degree of deviation of the resource consumption from the reference values is allowed. This maximum deviation depends on the general distribution of the resource consumption in the process plant around the baseline and has to be evaluated individually for every process plant. In Figure 4 an example of a baseline is shown in red and the boundaries of the resource consumptions that are considered to be resource efficient are indicated in green. These boundaries are calculated as percentages of deviation from the baseline value in order to consider the change of the general behavior of the process plant in these boundaries.

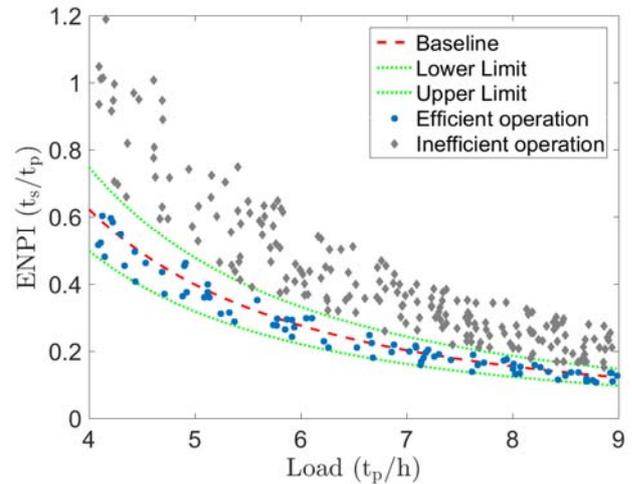


Figure 4: Baseline with process data and allowed deviations

The next step is to collect the available data from the process plant that is examined. To consider every possible effect that can influence the resource efficiency it is suggested to consider all of the data that is measured in the process plant and that is assumed to have an impact on the process plant. Furthermore, an extended period of time should be considered to increase the statistical relevance of the data-set. It is important that this timeframe covers data from resource efficient plant operation. An appropriate distribution of plant data can be seen in Figure 4. The whole operating window is covered with dense data and additionally sufficient data is present in the area of efficient operation. For each of these operation points all available measurements are acquired to enable a comparison to efficient operation points in the next steps.

After collecting the historical data the current operation point has to be added into the data-set. In general the estimation weather data is similar to the current operation point is not trivial because a lot of different parameters can change the plants performance. Using the baseline concept the current operation point can be easily introduced to the data-set via the uninfluenceable parameters that are used as inputs to the model. The EnPI is calculated for the current operation point and compared to the BDP of the inputs. The current operation point is indicated by the red point in Figure 5.

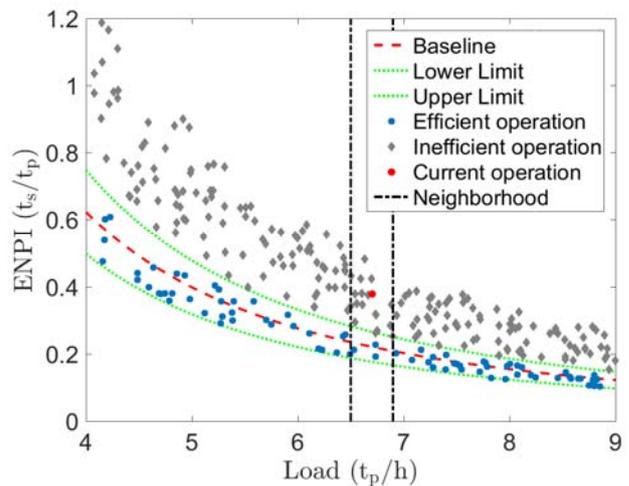


Figure 5: Baseline with current operation and a defined neighborhood

It is followed by the definition of a neighborhood for the current operation point. This neighborhood is defined as an area with values that are similar to the current operation point. The similarity is determined by the inputs used in the black-box model. By allowing deviations from the values for the inputs of the current operation an area is defined which includes historical data points. The size of the neighborhood is defined by the amount of resource efficient data points, which are indicated as blue dots in Figure 5. This is the “golden batch” of the data and represents desirable operation conditions. To ensure that this “golden batch” contains enough data points to compare them to the current operation point a minimal number of operation points is implemented as a threshold. The boundaries of the neighborhood are increased until the threshold for the minimal number of efficient operation points is met and the maximal distance of the inputs to the current operation point is then used as a limit for the allowed deviation of the inputs in the neighborhood.

Because the measurement data in process plants is suspect to a lot of changes and some parameters (e.g. ambient temperature) underlie natural variability, a direct comparison of the current parameters to the mean of the “golden batch” is highly susceptible to errors. To circumvent this limitation of the comparison and to determine which measurements are significantly different and likely to have an impact on the efficiency of the process the data has to be further processed. In order to take the effect of the variance of the data into consideration, the “golden batch” can be standardized and normalized by using the mean of the data (μ) and the standard deviation (σ) described by the following equations:

$$\mu_j = \frac{1}{N} \sum_{i=1}^N A_{i,j} \quad (1)$$

$$\sigma_j = \sqrt{\frac{1}{N-1} \sum_{i=1}^N |A_{i,j} - \mu_j|^2}, \quad (2)$$

where N describes the number of operation points in the “golden batch”, $A_{i,j}$ is the value of the measurement j for the operation point i , where the subscripted j indicates the different measurements (e.g. temperature sensor, pressure sensor...) of the process plant that is examined.

After this procedure the deviation $\delta_{k,j}$ in standard deviations of the “golden batch” can be calculated for all measurements j for the current operation point using the following equation:

$$\delta_{k,j} = \frac{A_k - \mu_j}{\sigma_j} \quad (3)$$

Where k indicates the current operation point. The deviation of each measurement can then be evaluated and has to be analyzed to estimate if the deviation of the resource consumption can be caused by the difference of this measurement for the current operation in comparison with the resource efficient operation points. Depending on the complexity of a process plant or a unit

operation it can have hundreds of measurements. This can be confusing and time consuming if all of them have to be analyzed at once.

To reduce the output of information to an amount that is more manageable a threshold is defined that has to be surpassed so that the deviation will be shown to the plant operator. This is based on the theory that deviations between measurements of the current operation point and the “golden values” are more likely to be the cause for different plant behavior when they are further away from each other. This can also be used to identify additional influence factors that should be used to model the resource efficiency of the plant more accurately if they are uninfluenceable by the plant operator or considered to be hard constraints that have to be maintained during the plant operation (e.g. minimal purity of the product). The analysis can further be facilitated by sorting the responses of the tool from the ones with the biggest deviation from the optimal values and showing them to the operator first. This can either be done by giving a graphical output in a monitoring tool that also indicates the optimal value for that measurement or by simply returning a message that a specific measurement deviates from the mean.

The last step is not automatic and has to be done by a person that is experienced with the plant and the physical principles behind it. The step involves the analysis whether or not the deviation is likely to be the cause for the different resource consumption. If an important influence factor has been identified the plant operator can adjust the current values to the mean values of the “golden batch” data. The mean values used to standardize the data can then be assumed to be a heuristic rule to give the plant operator a reference value for this specific measurement as to where it was in the past when the resource consumption was close to the baseline.

4.1.2 Divis approach to automatic rule extraction at INEOS in Cologne

The second approach used at INEOS in Cologne is based on the information about the Best Demonstrated Practice (BDP) that was already discussed in the previous section. The concept of

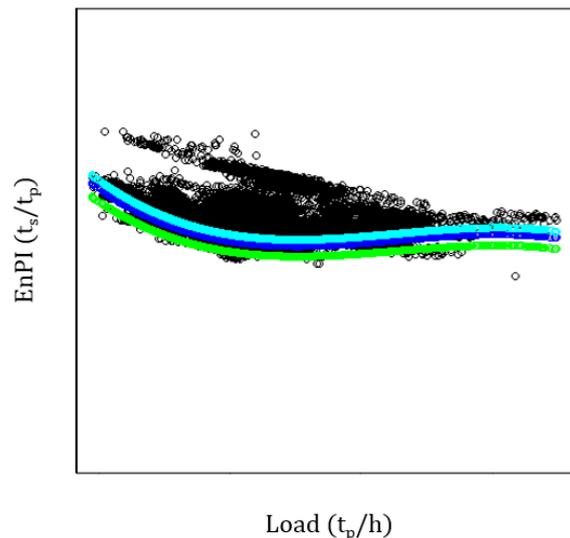


Figure 6: Classification of the data for a process unit into efficient and inefficient by using the BDP baseline

CART analysis presented in section 3.1 is used to identify the reasons for resource inefficient production of the unit operations in order to define heuristic rules that help to avoid the recurrent inefficient operation in the future. Similar to the previous method the implementation of the CART concept also benefits from the fact that the process industry collects lots of data that can be used. To

capture as much information as possible all available measurements of the unit operation are used for this analysis. Analogous to the concept described in section 4.1.1 a range of deviations from the baseline has to be allowed to classify the data into efficient and inefficient data points.

Figure 6 shows an example of the classification for a unit operation at INEOS in Cologne. The x-axis represents the load of the unit operation, which influences the resource consumption. The EnPI on the y-axis is the specific steam consumption. The dark blue line in the middle is the baseline for efficient production, the light blue line indicates the maximum allowed EnPI that is considered efficient and the green line indicates inefficient production respectively. Due to the relative simplicity of the mathematical formulation of the CART analysis it is possible to analyze the data of multiple years for a high amount of measurements in a short time. Longer timeframes that are analyzed in this way increase the probability to find operation points that were operated inefficiently so that they can be reduced to rules for plant operators.

5 Results and discussion

In this section the results of the methods presented in section 4.1.1 and section 4.1.2 are described and discussed. They are used on different unit operations at INEOS in Cologne and examples are shown to demonstrate the potential of these methods to increase the resource efficiency in the process plants solely by enhanced operation of the plants.

5.1 Discussion of the Root-Cause analysis

The Root-Cause analysis has been implemented in multiple plants at INEOS in Cologne and one case where potential savings were detected and finally realized will be discussed in this sub-section. The approach described in section 4.1.2 was used for an evaporator at INEOS in Cologne. A significant

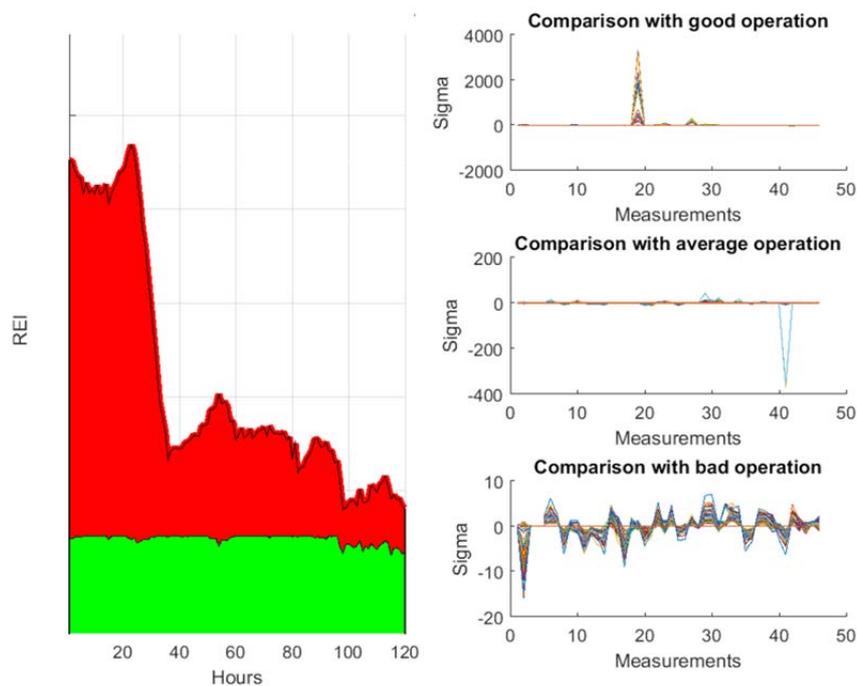


Figure 7: Implemented tool for automatic rule detection in an evaporator at INEOS in Cologne

deviation of the resource consumption to the baseline has been identified.

Figure 7 shows the implementation of the automatic rule extraction for the evaporator operation. The red line shows the current REI while the green line represents the baseline value for efficient production at the given plant conditions. It can be seen that a single measurement deviates approximately 2000 sigma from the mean value of good operation points. In this case this measurement was the reflux of the plant. It was significantly larger than the value it was compared to. After discussing this deviation with the plant personnel they adjusted the reflux to a more fitting value for the current operation conditions resulting in a steep decline of resource consumption between the 20th and the 40th hour in Figure 7. The value for the reflux that was proposed by the tool was then used by the plant personnel in a second step between the 80th and the 100th hour shown in Figure 7, reducing the resource consumption even further. Figure 8 shows the tool for automatic rule extraction after the adjustments for the current operation were implemented. It can be seen that the deviation of the current operation point to the good operation point is greatly reduced. The values for the deviation of each measurement between the new operation point and the mean of the

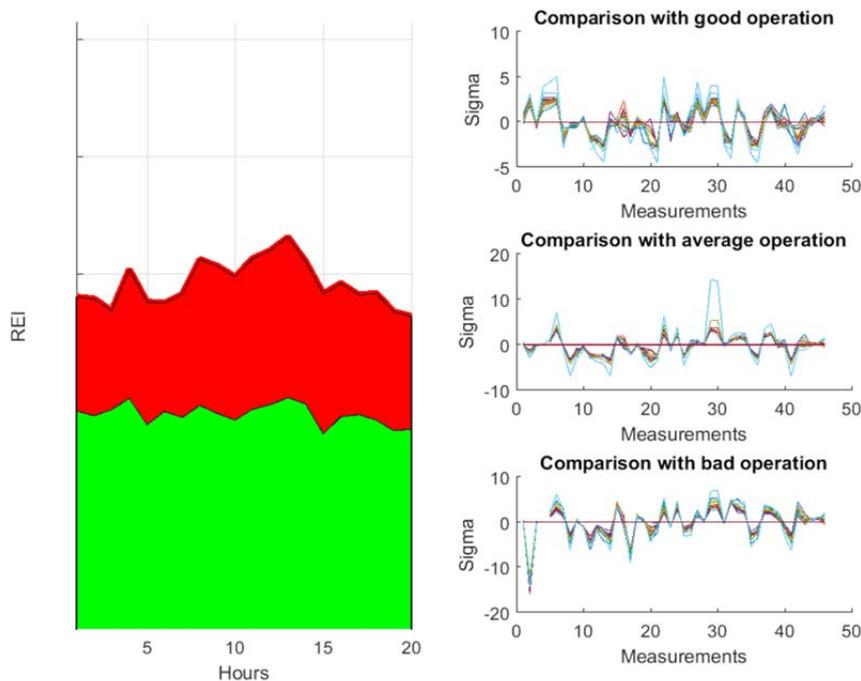


Figure 8: Demonstration of the implemented tool after the adjustments decreasing the resource consumption

resource efficient operation points are closer to each other and range around 5 sigma. The resource consumption in this case has been decreased by almost 40% by using one single hint. It is highly likely that the resource efficiency of the plants can be greatly increased by using a baseline tool for a plant and combining it with a tool that is able to determine measurements that are different.

5.2 Discussion of the CART-analysis

The resource efficiency of multiple process plants has been evaluated using the CART-analysis and the results are discussed in this section of the report. The example used to demonstrate this case is a distillation column.

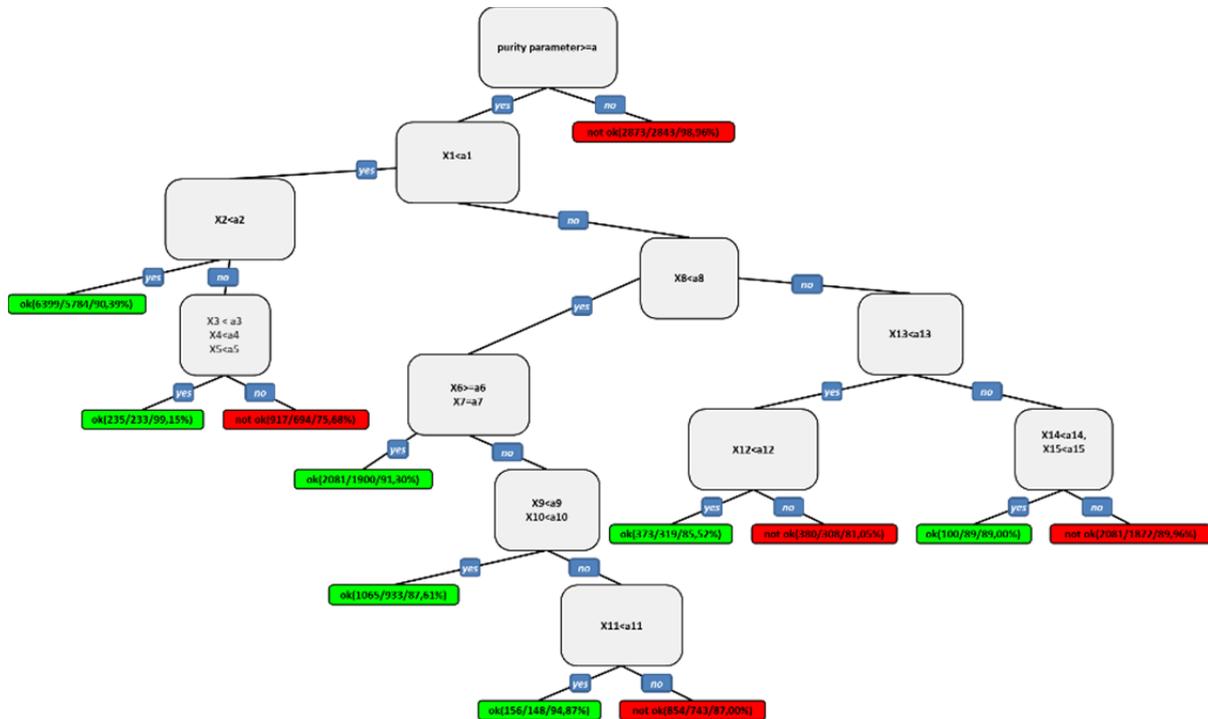


Figure 9: CART tree for a distillation column at INEOS in Cologne

The baseline for the BDP is used to indicate efficient operation and it is assumed that data 5% below and 3% above these values are still considered to be optimal. Historical data points are labeled as efficient if they fall into this range and the rest is labeled as inefficient. The goal was to find rules, which describe why the process is deviating from the optimal resource consumptions. Figure 9 shows the resulting CART regression tree with the corresponding rules that were used for the separation of the data. Additionally, the purity of each subset is indicated by the amount of values that fit into the assigned category and the total number of points in the node.

Looking at Figure 10 it can be seen that the first split, which is based on a purity measure, identifies almost 40% of the inefficient observations. One heuristic rule that could be implemented using these results could be: “never exceed this degree of purity”. It could be realized by sending a warning if the purity is approaching this value and should then be reevaluated by the plant operator. This classification and rule extraction approach can be applied to multiple areas with the goal to find logical rules, why observed data is structured in a specific way or is different from other data.

By using the rule that was derived for the distillation column, a timeframe has been identified in which the purity was very high for a prolonged time. Figure 11 shows a time-series for one impurity measurement of the distillation. The x-axis represents the time and the y-axis represents one measured impurity of the column. In the time between the two vertical red lines the impurity is significantly lower than it usually is.

Deliverable 2.6

Report on automatic rule extraction methods and results

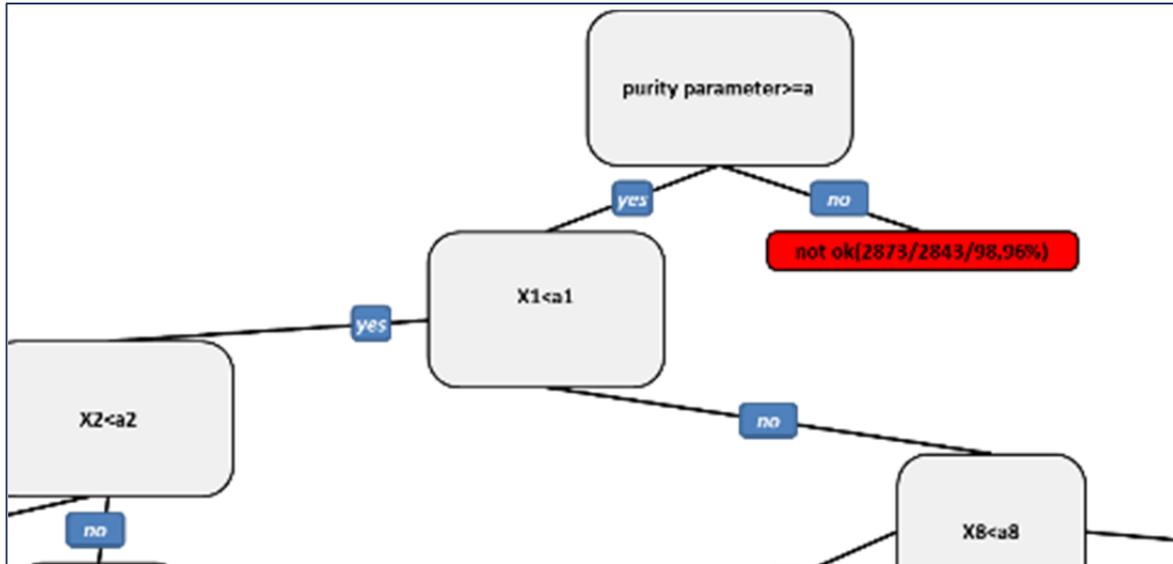


Figure 10: Upper part of the regression tree for the Best Demonstrated Practice of the resource consumption of a distillation column at INEOS in Cologne

This over-specification of the product caused an increase of 10% in the specific resource consumption when compared to the values obtained by the baseline model.

It should also be noted that the quality of the rules that are extracted using this method heavily relies on the purity of the resulting nodes. Only nodes with very high purity can be used for the automatic rule extraction. This is the reason why only a few rules could have been extracted for the plant operation even though the CART regression returns huge trees with different rules. Nevertheless, the implementation of the automatic rule extraction can help the plant operators to prevent inefficient operation by identifying bad operation from the past and notifying them, if the current operation is approaching these values.

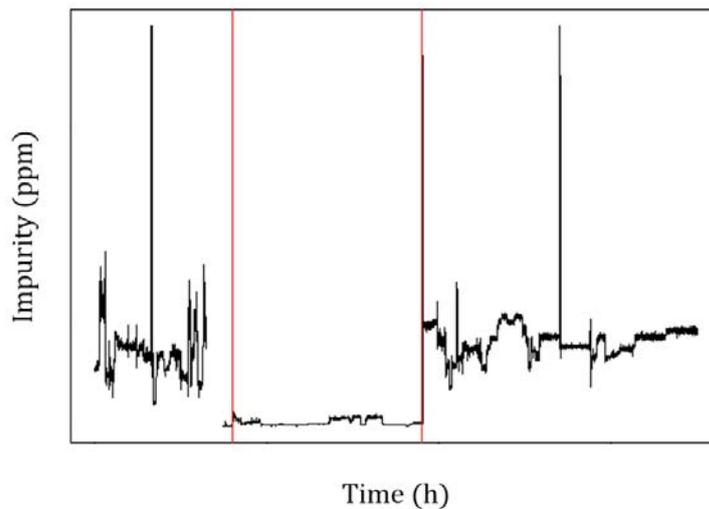


Figure 11: Time series for one impurity measurement of the distillation column

6 Summary and outlook

In this report two methods for the automatic rule extraction in chemical plants have been presented. Both of these concepts use historical data that is readily available in the process industry to analyze deviations between efficient operation and inefficient operation. They are most notably different in the approach they use. The root-cause analysis is comparing the current operation to efficient plant data to find differences in the measurements and point the plant operators to reference values for these measurements in order to increase the efficiency of the plant. The CART analysis compares measurement data of inefficient and efficient plant operation to each other and estimates which values are behaving differently in each of these sets.

The root-cause analysis can be used for live monitoring of the process and it could be shown that it helps the plant operators to achieve values closer to the baseline for resource efficient operation. The CART analysis results in general heuristic rules for plant operators that are taken from inefficient plant operation in the past. These can be used to establish boundaries for resource efficient operation.

It should be noted that both concepts do not result in the obsolescence of a plant operator. These concepts are developed to aid the operators to a more efficient operation by giving hints and relying on the fact that the plant operator is able to distinguish between useful and meaningless hints. Furthermore, the implementation of the described procedures always requires a human in the loop. The boundaries for efficient production have to be estimated and the results have to be evaluated by personnel with expert knowledge. Another restriction to the concepts that were developed is that the data points have to be classified as efficient and inefficient. In the described cases a baseline black-box model has been used for that purpose. However, even with all these constraints it could be shown that the implementation of these concepts increases the resource efficiency of the production plants.

Even though the resource consumption of process plants can already be reduced by these concepts there is further potential. To further automatize the whole rule extraction process the fluctuations of measurements in process plants has to be researched deeper in order to define more suitable thresholds for the deviations used in the root-cause analysis currently implemented rather arbitrarily by the model developer.

Another required step in both concepts is the definition of allowed deviations from the reference value given by the BDP baselines. This step can be further optimized by finding universally applicable methods that are able to accurately extract a range of resource efficient historical data.

Furthermore, the selection of measurements implemented in both of these concepts can be optimized by analyzing of the dependencies between each measurement and the corresponding resource consumption. An implementation of this step reduces the number of measurements and ultimately the amount of false positives (e.g. deviations in the wind measurements on a windy day) in both of these concepts.

However, in this report it could be shown that the implementation of the automatic rule extraction is a beneficial concept in order to reduce the resource consumption in chemical process plants.

7 References

B. Beisheim, K. Rahimi-Adli, S. Krämer, S. Engell. 2018. Energy performance Analysis of continuous processes using surrogate models. *Submitted to Energy*. 2018.

L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone. 1984. *CART: Classification and Regression Trees*. Belmont : Taylor & Francis Ltd, 1984. ISBN-10: 0412048418.